A NOVEL MULTIPLEX FOR SIMULTANEOUS AMPLIFICATION OF 20 Y CHROMOSOME STR MARKERS

John M. Butler[1*], Richard Schoske[1,2], Peter M. Vallone[1], Margaret C. Kline[1], Alan J. Redd[3], and Michael F. Hammer[3]

[1]*National Institute of Standards and Technology, Biotechnology Division, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899 USA*
[2]*American University, Department of Chemistry, Washington, DC 20016 USA*
[3]*University of Arizona, Division of Biotechnology, Tucson, AZ 85721*

ABSTRACT

A multiplex polymerase chain reaction (PCR) assay capable of simultaneously amplifying 20 Y-chromosome short tandem repeat (STR) markers has been developed to aid human identity testing and male population studies. These markers include all of the Y STRs that make up the "extended haplotype" used in Europe (DYS19, DYS385, DYS389I/II, DYS390, DYS391, DYS392, DYS393, and YCAII) plus additional polymorphic Y STRs (DYS437, DYS438, DYS439, DYS447, DYS448, DYS388, DYS426, GATA A7.1, and GATA H4). Primers for the markers DYS385, DYS389, and YCAII target duplicated regions of the Y chromosome and thus can provide two polymorphic peaks for each respective primer set. This Y STR 20plex, which utilizes 34 different PCR primers, is the first to include a simultaneous amplification of all the markers within the European "minimal" and "extended" haplotypes. Relative primer positions are compared between the newly developed primers described here and previously published ones. Efforts were made to avoid X chromosome homology in the primer design as well as close packing of PCR product size ranges in order to keep all alleles less than 350 bp through careful examination of known allele ranges. Haplotype comparisons between the 20plex and a commercially available kit found excellent agreement across the 76 samples in the Y chromosome consortium panel.

Keywords: Y chromosome; short tandem repeat DNA typing; multiplex PCR amplification; DYS19; DYS385; DYS388; DYS389; DYS390; DYS391; DYS392; DYS393; DYS426; DYS437; DYS438; DYS439; DYS447; DYS448; Y GATA A7.1; Y GATA H4; YCAII

*Corresponding author: Tel: 301-975-4049; Fax: 301-975-8505; Email: john.butler@nist.gov

1. INTRODUCTION

The Y chromosome is becoming a useful tool for tracing human evolution through male lineages [1] as well as application in a variety of forensic situations [2] including those involving evidence from sexual assault cases containing a mixture of male and female DNA [3,4]. Using Y-chromosome specific methods can improve the chances of detecting low levels of male DNA in a high background of female DNA. However, until recently use of the Y chromosome for forensic purposes was restricted by a lack of polymorphic markers. Since the Y chromosome is passed down from father to son without any recombination, individual markers cannot be combined using the product rule as is commonly done with autosomal STRs occurring on separate chromosomes [5]. Ideally, a forensic Y chromosome haplotype should include as many polymorphic loci as possible to improve the chance of excluding individuals (or male lineages) who did not commit the crime. In order to obtain a high level of discrimination either a large number of Y STRs can be run one at a time or combined into a multiplex.

Within the last five years, a number of Y STR multiplex assays have been developed [3-4, 6-8]. These multiplexes typically involve the simultaneous amplification of six or fewer loci. The European Y chromosome typing community has established a "minimal haplotype" and an "extended haplotype" for inclusion of common loci into a central DNA database (see http://ystr.charite.de). These haplotypes consist of results from the following Y STR markers: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, and YCAII [9]. The combination of these Y STRs can distinguish approximately 74 % to 90 % of male individuals in various local European populations [2]. The minimal haplotype, the results of which have been accepted for

court use in Europe [10], is most commonly used and is typically obtained via two or three separate multiplex amplifications. In this study, a Y STR 20plex capable of the simultaneous amplification of 20 polymorphic Y chromosome-specific PCR products is introduced. This 20plex includes all of the markers in the European extended haplotype [9] and also contains the trinucleotide loci DYS388 [2] and DYS426 [11], the tetranucleotide loci DYS437 [12,13], DYS439 [12,13], GATA A7.1 (DYS460 [38]) and H4 [14,15], the pentanucleotide loci DYS438 [12,13] and DYS447 [16], and the hexanucleotide marker DYS448 [16].

2. MATERIALS AND METHODS*

2.1 DNA Samples

The primary set of DNA samples examined in this report came from the Y chromosome consortium (YCC) panel. The YCC panel consists of 74 male samples and 2 female samples composed of individuals from diverse world populations and originally established as cell lines in the Laboratory of Human Genetics at the New York Blood Center (see http://www.arl.arizona.edu/lmse/ycc/ycc1.html). The YCC samples were provided to the National Institute of Standards and Technology (NIST) from the University of Arizona as extracted DNA that had been previously quantified.

Additional studies were performed with 50 anonymous human blood samples purchased from a commercial blood bank (Millennium Biotech, Inc., Ft. Lauderdale, FL). These samples were used in our multiplex optimization, were approved for use through

* Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology or the US Department of Defense nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

an institutional review board at NIST, and were extracted using conventional organic extraction methods [17].

2.2 Development of reference sequences

Sequences for each locus were obtained from GenBank® (www.ncbi.nlm.nih.gov) using a standard nucleotide BLAST (Basic Local Alignment Search Tool) search. The previously published Y STR marker primers were originally chosen as the query sequences for each Y STR locus. If a query sequence returned multiple GenBank accession numbers, then the most recent entry or the sequence with the larger flanking regions to the repeat was used for further evaluation including sequence alignments. Sequence alignments of the STR repeat and 200 base pair (bp) to 300 bp flanking regions for the various GenBank sequence entries were performed to aid detection of polymorphic nucleotides and regions where X chromosome homology was known to exist (see Results and Discussion). Sequence alignments were accomplished through the Baylor College of Medicine (BCM) search launcher located at http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html.

From these sequence alignment efforts, a consensus reference sequence was produced that originated from the GenBank® accession numbers listed for each Y STR marker in Table 1. The number of repeats contained within each reference sequence was then determined using ISFG guidelines [18] (see Table 1). It should be noted that the allele nomenclature for Y GATA H4 shown in Table 1 is different than previously published information [15] because our reverse primer is internal to some of the invariant repeats—a fact that may impact future database compatibility with different primer sets.

2.3 Multiplex primer design and quality control

The allele size ranges of the Y STR 20plex loci listed in Table 2 and illustrated in Figure 1 were defined through extensive literature searches and testing of the YCC panel [19]. A listing of Y chromosome literature examined in defining the known allele ranges may be found in STRBase [20]: http://www.cstl.nist.gov/biotech/strbase/y_strs.htm. A multiplex PCR design strategy was followed as outlined elsewhere (Schoske *et al., submitted*). Briefly, Y STR loci were selected and their allele ranges carefully defined based on the literature examinations. Loci were laid out in terms of possible PCR product size and dye color (see Figure 1). Then primers were designed with Primer 3 [21] using a defined PCR product size to match the required allele range on a consensus reference sequence (See 2.2)

Candidate primers were then screened for potential primer cross reactions using an algorithm previously described [22]. Primers were purchased from MWG Biotech (High Point, NC), Operon Technologies (Alameda, CA), or Applied Biosystems (Foster City, CA) and upon receipt were quality control tested prior to further use to confirm proper synthesis using methods previously described [23]. The primer sequences used in the Y STR 20plex are listed in Table 3.

2.4 Y STR 20plex PCR amplification conditions

The Y STR 20plex amplifications were performed in reaction volumes of 20 µL with 2 units of AmpliTaq Gold® DNA polymerase (Applied Biosystems), 10mM* Tris-HCl (pH 8.3), 50 mM KCl, 1.75 mM $MgCl_2$, 300 µm deoxynucleotide triphosphates

* The accepted SI unit of concentration, mol/L, has been represented by the symbol M in order to conform to the conventions of this journal.

(dNTPs: dATP, dCTP, dGTP, dTTP), 5% (v/v) glycerol, 4.4 µL of the Y-STR primer mix (see Table 3 for concentrations), and 2 µL DNA template. For most of the studies described here, the DNA template amount was 5 ng or 10 ng. Thermal cycling was conducted on a GeneAmp 9700 (Applied Biosystems) using the following conditions in 9600-emulation mode (i.e., ramp speeds of 1 ^{o}C/s):

95 ^{o}C for 10 min
28 cycles: 94 ^{o}C for 1 min, 55 ^{o}C for 1 min, 72 ^{o}C for 1 min
60 ^{o}C for 45 min
25 ^{o}C hold

2.5 Y-PLEX^{TM} 6 kit

The commercially available Y-PLEX^{TM} 6 kit (ReliaGene Technologies, New Orleans, LA) was used to amplify DYS19, DYS385, DYS389II, DYS390, DYS391, and DYS393. PCR amplification conditions were performed as recommended by the manufacturer in the Y-PLEX^{TM} 6 instruction manual version 1.0, except that reaction volumes were reduced in half.

2.6 Autosomal STR Amplification

The PowerPlex® 16 kit (Promega Corporation, Madison, WI) was used according to manufacturer's protocols and 2.5 ng of DNA template. The PowerPlex® 16 System amplifies the following autosomal STR loci: Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Penta D, CSF1PO, D16S539, D7S820, D13S317, D5S818 and the gender typing marker amelogenin.

2.7 Filtration of PCR Products

An aliquot of 5 µL of amplified PCR products from a subset of the Y STR 20plex reactions was filtered using Edge Gel Filtration Cartridges (Catalog # 42453), provided by Edge BioSystems (Gaithersburg, MD), according to the manufacturer's protocol. A 2 µL aliquot of the filtrate was taken and analyzed according to the protocol outlined in the next section. PCR products were filtered in order to remove free dye labels that can interfere with data interpretation and allele designations. This filtration is useful post-amplification because the dye-labeled Y STR 20plex primers were purchased without purification in order to reduce the cost of constructing this multiplex assay.

2.8 Detection and analysis of PCR products

Detection of the Y STR 20plex amplified products was accomplished with the ABI Prism® 3100 Genetic Analyzer 16 capillary array system (Applied Biosystems) following manufacturer's protocols using the G5 matrix filter to detect the five dyes 6FAM™ (blue), VIC™ (green), NED™ (yellow), PET™ (red), and LIZ™ (orange) available from Applied Biosystems. Prior to running any samples, a spectral matrix was established using matrix standard set DS-33 (Applied Biosystems, P/N 4323016). A four dye matrix was also established under the Z filter with the dyes 6FAM (blue), VIC (green), NED (yellow), and ROX (red) using matrix standard set DS-30 (P/N 4316100) and substituting the VIC matrix standard (P/N 4323022) for HEX. Samples were prepared with 18.6 µL Hi-Di™ formamide (Applied Biosystems, P/N 4311320), 0.4 µL GS500 LIZ size standard (P/N 4322682) or 0.75 µL GS500 ROX (P/N 401734), and with either 1 µL PCR product or 2 µL of filtered PCR product (see 2.6). Samples were then

injected onto the 16-capillary array for 10 s at 3,000 V. Separations were performed at 15,000 V for 44 min with a run temperature of 60 °C using the 3100 POP™-4 sieving polymer (Applied Biosystems, P/N 4316355), 1X Genetic Analyzer Buffer with EDTA (P/N 402824), and a 36 cm array (P/N 4315931). Following data collection, samples were analyzed with Genescan® 3.7 (for Windows NT, Applied Biosystems), and allele designations were determined using Genotyper® 3.7 (Applied Biosystems) based on a genotyping macro modified from the Kazam macro used in the ABI AmpFlSTR® Identifiler™ kit. Allele designations were made based on sizing bin windows of up to +/- 1.50 bp rather than by comparison to allelic ladders.

The separation and detection of PowerPlex® 16 kit and Y-PLEX™ 6 kit generated PCR products were accomplished with an ABI Prism® 310 Genetic Analyzer (Applied Biosystems) using filter set A. An appropriate matrix was established with matrix standards for the four dyes FAM, JOE, TAMRA, and ROX (Applied Biosystems, P/N 401546) or FL, JOE, TMR, and CXR (Promega, Cat # DG2860). Each sample was prepared by adding 1 µL PCR product to 20 µL of deionized formamide containing 0.75 µL GS500 ROX size standard or 1 µL ILS600 CXR (Promega, Cat #DG2611). Samples were injected for 5 s at 15,000 V and separated at 15,000 V for 26 min with a run temperature of 60 °C using 310 Genetic Analyzer POP™-4 (P/N 402838), 1X Genetic Analyzer Buffer with EDTA, and a 47 cm x 50 µm capillary (J & W Scientific, Folsom, CA). Following data collection, samples were analyzed with Genescan® 3.1 (for Macintosh, Applied Biosystems) and allele designations were determined by comparison to allelic ladders using Genotyper® 2.5 (Applied Biosystems) and the Y-PLEX™ 6 310

*v*3.0 Genotyping Template provided by ReliaGene with their kit. PowerPlex® 16 samples were analyzed with the kit allelic ladders and PowerTyper™ 16 macro (Promega).

3. RESULTS AND DISCUSSION

3.1 Selection of loci

In deciding which Y STR loci should be included in our megaplex, we examined the literature to find the most commonly used markers and found that the European community has settled upon a set of markers commonly referred to as the "minimal haplotype" and the "extended haplotype" [9]. The minimal haplotype includes the Y STR markers DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, and DYS385. The extended haplotype includes all of these loci plus the highly polymorphic dinucleotide repeat YCAII [24,25]. At the time we first started examining Y STRs several years ago there were not many more polymorphic Y STRs available. The number of loci available grew with the addition of DYS434, DYS435, DYS436, DYS437, DYS438, and DYS439 by Ayub *et al.* [12] and GATA A7.1 (DYS460), A7.2 (DYS461), H4, A8, A10, and C4 by White *et al.* [14]. Thus, we decided to include in our multiplex the most polymorphic markers from these two references and settled on DYS437, DYS438, DYS439 and GATA A7.1 (DYS460) and H4. These five markers were part of a Y STR 10plex described previously [26]. The trinucleotide repeat loci DYS388 [2] and DYS426 [11] were also included, as they have been used in previous Y chromosome studies. Lastly, DYS447 and DYS448, which are part of the newly discovered Y STR markers reported in the accompanying paper [16], have been included in the multiplex described here.

3.2 Primer design

In order to get as many Y chromosome markers into the multiplex, we followed a design strategy to pack as many loci as possible into each dye color. This approach requires prior knowledge of all previously observed alleles to minimize the chance for overlap between PCR products amplified with the same fluorescent dye if new alleles are discovered after the multiplex has been developed. Typically, a spread of 10 base pairs was left between known alleles for loci labeled with the same dye after the maximum and minimum repeat sizes were determined through extensive evaluation of allele ranges for each Y STR marker.

In order to increase the number of markers that could be analyzed and keep the PCR product sizes relatively small, we took advantage of the newly available five-dye capabilities of ABI PRISM® instruments. The use of an extra dye color for labeling PCR products increases the number of potential markers that may be amplified and distinguished in a multiplex amplification. We used the dyes 6FAM (blue), VIC (green), NED (yellow), and PET (red), which are all commercially available, to label the PCR products while a fifth dye LIZ (orange) was used on the internal size standard. It is also possible to remove the loci labeled with the red dye and run a subsection of the multiplex described here on four-dye detection platforms using the red dye lane for the internal size standard. To simplify primer purchasing, only the forward primers were labeled at the 5' end with 6FAM, VIC, NED, or PET (Table 3).

A G nucleotide tail was added to the 5'end of the reverse primer to promote non-template addition of the PCR products [27,28]. With the addition of a non-template base

plus the G 5'-tail, the labeled, fully adenylated PCR product is two bases longer than an amplicon predicted on a reference sequence from the primer positions alone. Without the addition of the G 5' tail the fully adenylated PCR product is one base pair longer than an amplicon predicted on a reference sequence from the primer position alone. The new size ranges listed in Table 2 take this information into account.

Previously published primer sequences were used for DYS385 [29], DYS390 [2], DYS393 [2], DYS426 [11], and DYS437 [12]. Due to the fact that the DYS385 a/b primers target a duplicated region of the Y chromosome, the resulting Y STR haplotype is highly polymorphic. For this reason, we built this Y STR 20plex around primers proven to successfully amplify DYS385.

Primers were redesigned for most of the other loci for three reasons. First, PCR product size ranges had to be adjusted from previous studies in order to optimize their sizes in each dye color (i.e., to prevent overlapping between loci that complicates data interpretation or leaving large gaps that essentially waste space). PCR product sizes can most easily be adjusted by moving the primer positions in the flanking regions surrounding the Y STR repeat segment.

Second, primers need to have similar annealing characteristics in order to generate a balanced yield from all of the PCR products during a simultaneous amplification. Prior to this work, most Y STR loci were initially designed to amplify the markers in a singleplex fashion or limited multiplexes and therefore were not selected to work efficiently together. For example, the previously used YCAII primers [24], 5-TATCGATGTAATGTTATATTA-3 and 5'-TATATTAAATAGAAGTAGTGA-3', have a predicted melting temperature (T_m) of 43.0 ^{o}C and 39.2 ^{o}C, respectively. The new primers

for YCAII designed here to work in the Y STR 20plex have a T_m of 59.3 ^oC and 57.2 ^oC (Table 3). The higher T_ms for these primers dramatically improves their amplification efficiency particularly in a multiplex PCR environment.

Third, the latest knowledge about X chromosome homology or polymorphic nucleotides in primer binding regions was applied to avoid regions that would yield signal in female DNA samples or impact PCR amplification. The best example of an X homolog problem is DYS391 [30]. Figure 2 illustrates the individual sequences for X and Y homologs of the DYS391 locus along with a sequence alignment of these two regions. The new primers described in this work exploit sequence differences in order to maximize primer binding potential to only the Y homolog in order to make the amplification Y chromosome specific (Figure 2). While Gusmao and coworkers [30] targeted four sequence differences in their improved reverse primer for DYS391, our newly designed primers target one sequence difference for the forward primer and eleven sequence differences in the reverse primer to improve Y allele specific amplification for DYS391.

3.3 Initial testing

All primer pairs were initially tested under identical amplification conditions including the same DNA template concentration. Following the successful individual runs of each primer pair they were combined into a multiplex. If amplification of a particular locus was poor, the primer concentration for that locus was increased and the test was repeated. Unlike previously published multiplex PCR primer mixture design protocols [31,32] in which many other experimental conditions, such as $MgCl_2$

concentration, *Taq* concentration, dNTP concentration, buffer concentration, annealing temperature adjustments, and PCR cycle numbers were modified to achieve better amplification, only the primer concentrations were varied and needed to be slightly adjusted with our approach because the primers were designed initially to work together.

After running multiple male test samples, the primer concentrations were adjusted in order to achieve an improved balance in the PCR products generated by the multiplex. The primer concentrations used in this study are in some cases higher than those in other multiplex work [6,7]. Primer concentrations were used as the primary means to increase the amount of PCR product balance. The higher primer concentration permits for more robust amplification without increasing items such as cycle number, $MgCl_2$ concentration, buffer concentrations, annealing temperatures and *Taq* concentrations. For the initial construction of the multiplex all of the primer pairs were combined at a concentration of 1.0 µM. If one PCR product was higher in peak height relative to the other amplicons in the multiplex, the appropriate primer pair concentrations was decreased to try to generate a more balanced yield between the various PCR products. Relative peak heights from the amplicons in the Y STR 20plex were used to estimate the needed primer concentrations adjustments in order to improve the amplification balance. After several iterations and adjustments of primers concentrations, a final primer mix was decided upon, which generates a set of fairly balanced PCR products (Figures 3 and 4).

3.4 New primer positions relative to previous sets

Table 4 lists the relative primer positions between the new primers reported here and previously published primer sets that are widely used, particularly in the European

community. This primer position information is provided to help in deciphering potential problem areas if differences in Y chromosome typing are observed in the future. While we have yet to see any problems with our new primer sets in terms of allele dropout (see below), a very real concern exists that null alleles can potentially arise when different primer sets are used. Differences in results between samples from the same source but amplified with different primer sets could impact DNA databasing efforts [5]. In our initial primer design, we made every effort possible to avoid known polymorphic nucleotides in the primer binding sites through sequence alignments of multiple GenBank® entries.

During the development of new primers for the locus DYS19, a discovery was made of a duplication of the flanking regions for this locus on another portion of the Y chromosome. The original primers for DYS19 [1,33] have a low GC content and a T_m that is too low for optimal performance in our multiplex. New primers were designed and tested only to discover that a second region of the Y chromosome was also being amplified. The PCR product produced was not polymorphic in over 100 male samples tested and measures 187 bp using the conditions described in the materials and methods section. A detailed sequence search through GenBank® entries located a second significant match to the new DYS19 primers. A subsequent sequence alignment revealed that almost the entire flanking region of the original DYS19 locus has been duplicated elsewhere on the Y chromosome and that primers and PCR conditions have to be carefully selected to avoid this duplication (Schoske *et al.*, in preparation). The DYS19 primers listed in Table 3 amplify only the DYS19 locus by exploiting sequence

differences between the DYS19 flanking region and the duplicated portion of the Y chromosome.

It is interesting to note that in the course of examining newly available human genome sequence information, we discovered that the DYS385 duplications (i.e., DYS385 a/b) are separated by 40,775 bp and the YCAII duplications (i.e, YCAII a/b) are separated by 885,555 bp along the long arm of the Y chromosome. Thus, bioinformatics has the potential to greatly enhance our knowledge of the Y chromosome as well as aid in design and optimization of new assays.

3.5 Comparison of Y STR 20plex typing information to the Y-PLEX™ 6 kit

To compare the concordance of allele designations with our new assay and other available primer sets, haplotypes were generated on the same set of male DNA samples (Table 5) using the Y STR 20plex and the Y-PLEX™ 6 kit available from ReliaGene. These two assays have in common seven typing results from six different Y STR markers that may be compared. The Y STR markers in common between the two assays are DYS19, DYS393, DYS389II, DYS390, DYS391, and DYS385I/II.

Figure 3A depicts the result from a male DNA sample run on an ABI 3100 with the Y STR 20plex assay while Figure 3B contains the result from the same male sample run on an ABI 310 with the Y-PLEX™ 6 kit. The allele calls are completely concordant even though they were generated on different instruments, by different primers, and by different allele calling approaches. Allele calling with the Y-PLEX™ 6 assay was done by comparison of allele sizes to sequenced allelic ladders provided in the kit. On the other hand, the Y STR 20plex allele calls were made with a Genotyper macro that uses

categories with allele bin windows of +/- 0.75 bp to +/- 1.50 bp. The bin windows were established based on sequence information from a single allele and the allele sizing data provided by an internal sizing standard (e.g., GS500 LIZ).

A comparison of Y chromosome haplotypes for the 74 male samples in the YCC panel found that out of 1036 possible allele designations there was complete concordance between the two approaches with the exception of 3 calls. The alleles that were the subject of the discrepant calls were all deemed "off-ladder" alleles by the Y-PLEX^{TM} 6 kit while the Y STR 20plex assay identified them as full repeats. Thus, the Y-PLEX^{TM} 6 kit produced a call of 9.3 at the DYS391 locus for sample YCC19 while the 20plex result was reported as allele 10. The other two samples in question (DYS385 locus in YCC 49 and 50) contained a 10.3 allele when compared to the Y-PLEX^{TM} 6 allelic ladder but were deemed 11 repeats in the 20plex assay. Amplicon sizes for the DYS391 and DYS385 loci generated by the Y STR 20plex are smaller than those generated by the Y-PLEX^{TM} 6 kit (see Figure 3). Thus, it is possible that these microvariants could be due to deletions that exist in regions lying outside of the Y STR 20plex primer binding sites and therefore not be detected with the new primer sets [39]. Of course, it is also possible that because the Y STR 20plex assay does not currently utilize allelic ladders and the bin windows are too large to confidently distinguish an allele containing a single base deletion from one with a full repeat, the three alleles in question were misclassified as the nearest full repeat instead of the variant. We plan to sequence the affected samples in order to ascertain the exact cause of the discrepancies.

The only case of allele dropout observed in our analysis of the YCC panel was at DYS439 with samples YCC29 and 30, which are both !Kung individuals that are

probably male cousins as they share alleles at 9 autosomal STRs (see Table 6). A different DYS439 primer pair, where the primers were completely non-overlapping from the ones in Table 3, failed to amplify these two samples suggesting that this portion of the Y chromosome may be deleted or there is a polymorphism in the primer binding regions for these samples (data not shown).

Table 5 contains a complete list of allele calls obtained with the 20plex assay for the 76 samples contained in the YCC panel. Samples YCC1 and YCC54 are female and did not generate any allele calls when amplified with the Y STR 20plex primer set using a DNA template amount of 10 ng. Sequencing of alleles from the loci amplified in the Y STR 20plex has been performed to verify the exact number of repeats.

The YCC panel is a rich source of diverse alleles for Y STR markers since many of these samples come from isolated African, South American, or Siberian populations [34]. In the six loci tested as part of the Y-PLEX^{TM} 6 kit, a total of 17 "off-ladder" alleles were observed including the microvariants at DYS391 and DYS385 mentioned above and one triallelic pattern observed at DYS385 in the YCC 22 sample. YCC samples work well for assay development but not for drawing thorough population genetics conclusions because there are not enough samples from any one population. In addition, during development of the YCC cell line panel, samples were not screened prior to collection to eliminate individuals from the same male lineage. Hence, six pairs of YCC samples are likely from the same or similar male lineages and cannot be distinguished from one another using the haplotypes generated by the Y STR 20plex. These pairs are YCC6/7, YCC 12/13, YCC 15/16, YCC 29/30, YCC 49/50, and YCC 8/37. Autosomal

STR results from 15 STR loci included a commercial STR multiplex verified that the samples in these pairs are not from the same individuals (Table 6).

Perhaps it is worth pointing out that the use of only the extended haplotype rather than the full 20plex would have done just as well at separating the various haplotypes from one another within the YCC panel. The availability of more Y STR markers in this particular case did not improve the power of discrimination although it likely will in other population data sets because of the additional Y STRs [38]. However, unlike the minimal and extended haplotypes, which previously required two to three amplifications with smaller multiplexes to obtain results [2,38], the Y STR 20plex can achieve a high power of discrimination in a single amplification.

If a laboratory, perhaps one performing paternity testing, desires to only run the 11 loci in the "extended haplotype" because of concerns with possible mutations being seen with the use of additional loci, then a subset of the 20plex can easily be made by selecting the desired loci from the primers listed in Table 3. Figure 4 demonstrates that the same typing results may be obtained by simply eliminating primers for the nine non-"extended haplotype" Y STR loci. The existence of a large multiplex now opens new possibilities for studying which Y STRs have the most significant impact on a particular population by enabling more rapid collection of the data from population samples. In addition, a combination of loci with a lower degree of stutter may be used to replace the highly polymorphic but stutter-prone dinucleotide locus YCAII. In this manner, the most polymorphic loci with the fewest artifacts could be applied to casework samples to simplify interpretation of sexual assault mixtures.

3.6 Initial Sensitivity and Mixture Studies

Some preliminary experiments were performed to explore the use of this Y STR 20plex with male-female mixtures as are commonly encountered in forensic casework. Initial sensitivity studies appear to indicate that the male component can be detected with as little as 250 pg of DNA using the standard 28-cycle PCR protocol. However, the presence of free dye from unpurified primers (i.e., dye blobs) in these electropherograms makes the automated calling of some alleles problematic. Future work will include testing whether or not sensitivity can be extended to lower levels of male DNA template by using primers that have been HPLC purified to remove all free dye impurities.

For the male/female DNA mixtures it was possible to detect the male component up to the highest tested ratio of 1:150 (data not shown). However, there appeared to be some interference with PCR products generated from female DNA at levels higher than approximately 40 ng. It is possible that these problems arise from the DYS385 [35,36] and DYS393 [37] primers that were kept consistent with previous work where some of the same problems were reported. For the male/male mixtures we were able to detect the male component up to the tested ratio of 1:10 (data not shown).

4. IMPLICATIONS AND CONCLUSIONS

A novel Y chromosome assay has been developed that offers a potential increased power of discrimination compared to previous Y STR multiplex assays. The ability to obtain information from 20 different sites along the Y chromosome in a single amplification should greatly speed database development and aid forensic investigations. While this report lays out the 20plex primers and PCR conditions, a great deal of work is

still required before the multiplex can be used routinely in casework. Population studies and evaluation of haplotype discrimination power will need to be examined in more detail as will the performance of the multiplex with casework samples. Allelic ladders need to be created for all of the loci in order to meet the recent International Society of Forensic Genetics recommendations on forensic analysis using Y-chromosome STRs [18].

The publication of the YCC panel haplotype data reported here will help standardize results obtained when different primer sets are used in the future. A Standard Reference Material (SRM) is under development in our lab at NIST that will aid in future comparisons of different primer sets for commonly used Y STR markers. SRM 2395 should be available in late 2002 and will include 5 male samples and 1 female sample that have been sequenced at the 20 loci described here. Laboratories wishing to verify that their assays were run properly *with any primer set* can use these reference materials.

In summary, the design, optimization, and testing of a multiplex mixture through careful primer design resulted in the construction of a multiplex that can simultaneously amplify 20 Y STR markers. In the near future the Y STR 20plex described here could become useful to the forensic and anthropological DNA testing communities because it encompasses all of the markers contained within the European 'minimal' and 'extended' haplotypes as well as 9 additional polymorphic Y STR markers. Using the Y STR 20plex in forensic casework could make the exclusions of suspects less ambiguous because of the inclusion of these additional Y STR markers. We anticipate that high-throughput databasing of Y STR population data will be greatly enhanced by the availability of this Y STR 20plex and future multiplexes under development.

ACKNOWLEDGEMENTS

This work was funded by National Institute of Justice research grants 1997-LB-VX-0003 to John Butler through the NIST Office of Law Enforcement Standards and 2000-IJ-CX-K006 to Michael Hammer. Richard Schoske is a PhD candidate in chemistry at American University under Professor Jim Girard and is funded by the United States Air Force, through the Air Force Institute of Technology. The technical assistance of Al Agellon, Veronica Kearney, and Tatiana Karafet from the University of Arizona is greatly appreciated in the identification of DYS447 and DYS448. The authors also thank Mecki Prinz from the New York City Office of the Chief Medical Examiner for helpful discussions as this multiplex was being designed.

REFERENCES

[1] M.A. Jobling, and C. Tyler-Smith, Fathers and sons: the Y chromosome and human evolution, Trends Genet. 11 (1995) 449-456

[2] M. Kayser, A.Caglia, D.Corach, N.Fretwell, C.Gehrig, G.Graziosi, F.Heidorn, S.Herrmann, B.Herzog, M.Hidding, K.Honda, M.Jobling, M.Krawczak, K.Leim, S.Meuser, E.Meyer, W.Oesterreich, A.Pandya, W.Parson, G.Penacino, A.Perez-Lezaun, A.Piccinini, M.Prinz, C.Schmitt, L.Roewer, Evaluation of Y-chromosomal STRs: a multicenter study, Int.J.Legal Med 110 (1997) 125-129.

[3] M. Prinz, K. Boll, H. Baum, B. Shaler, Multiplexing of Y chromosome specific STRs and performance for mixed samples, Forensic Sci. Int. 85 (1997) 209-218.

[4] M. Prinz, A. Ishii, A. Coleman, H.J. Baum, R.C. Shaler, Validation and casework application of a Y chromosome specific STR multiplex, Forensic Sci.Int. 120 (2001) 177-188.

[5] J.M. Butler, Forensic DNA Typing, Academic Press, San Diego, CA, 2001.

[6] A.J. Redd, S.L. Clifford, M. Stoneking, Multiplex DNA typing of short-tandem-repeat loci on the Y chromosome, Biol.Chem. 378 (1997) 923-927.

[7] L. Gusmao, A. Gonzalez-Neira, C. Pestoni, M. Brion, M.V. Lareu, A. Carracedo, Robustness of the Y STRs DYS19, DYS389 I and II, DYS390 and DYS393: optimization of a PCR pentaplex, Forensic Sci.Int. 106 (1999) 163-172.

[8] A. Carracedo, A. Beckmann, A. Bengs, B. Brinkmann, A. Caglia, C. Capelli, P. Gill, L. Gusmao, C. Hagelberg, C. Hohoff, B. Hoste, A. Kihlgren, A. Kloosterman, D.B. Myhre, N. Morling, G. O'Donnell, W. Parson, C. Phillips, M. Pouwels, R. Scheithauer, H. Schmitter, P.M. Schneider, J. Schumm, I. Skitsa, B. Stradmann-Bellinghausen, M. Stuart, C.D. Syndercombe, C. Vide, Results of a collaborative study of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format, Forensic Sci.Int. 119 (2001) 28-41.

[9] L.Roewer, M.Krawczak, S.Willuweit, M.Nagy , C.Alves, A.Amorim, K.Anslinger, C.Augustin, A.Betz, E.Bosch, A.Caglia, A.Carracedo, D.Corach, A.Dekairelle, T.Dobosz, B.M.Dupuy, S.Furedi, C.Gehrig, L.Gusmao, J.Henke, L.Henke, M.Hidding, C.Hohoff, B.Hoste, M.A.Jobling, H.J.Kargel, P.de Knijff, R.Lessig, E.Liebeherr, M.Lorente, B.Martinez-Jarreta, P.Nievas, M.Nowak, W.Parson, V.L.Pascali, G.Penacino, R.Ploski, B.Rolf, A.Sala, U.Schmidt, C.Schmitt, P.M.Schneider, R.Szibor, J.Teifel-Greding, M.Kayser, Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes, Forensic Sci. Int. 118 (2001) 106-113.; http://ystr.charite.de

[10] V.L. Pascali, M. Dobosz, B. Brinkmann, Coordinating Y-chromosomal STR research for the courts, Int.J.Legal Med. 112 (1999) 1.

[11] M.A.Jobling, V.Samara, A.Pandya, N.Fretwell, B.Bernasconi, R.J.Mitchell, T.Gerelsaikhan, B.Dashnyam, A.Sajantila, P.J.Salo, Y.Nakahori, C.M.Disteche, K.Thangaraj, L.Singh, M.H.Crawford, C.Tyler-Smith, Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males, Hum. Mol. Genet. 5 (1996) 1767-1775.

[12] Q. Ayub, A. Mohyuddin, R. Qamar, K. Mazhar, T. Zerjal, S.Q. Mehdi, C. Tyler-Smith, Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information, Nucleic Acids Res. 28 (2000) e8.

[13] P. Grignani, G. Peloso, P. Fattorini, C. Previdere, Highly informative Y-chromosomal haplotypes by the addition of three new STRs DYS437, DYS438 and DYS439, Int.J.Legal Med. 114 (2000) 125-129.

[14] P.S. White, O.L. Tatum, L.L. Deaven, J.L. Longmire, New, male-specific microsatellite markers from the human Y chromosome, Genomics, 57. (1999) 433-437.

[15] A. Gonzalez-Neira, M. Elmoznino, M.V. Lareu, P. Sanchez-Diz, L. Gusmao, M. Prinz, A. Carracedo, Sequence structure of 12 novel Y chromosome microsatellites and PCR amplification strategies, Forensic Sci. Int. 122 (2001) 19-26.

[16] A.J. Redd, A.B. Agellon, V.A. Kearney, T. Karafet, P. de Knjiff, H. Park, J.M. Butler, M.F. Hammer, Forensic value of fourteen novel STRs on the human Y chromosome, Forensic Sci. Int. (2002), submitted.

[17] J. Sambrook, E.F. Fritsch, and T. Maniatis. Molecular Cloning: A Laboratory Manual, Vols. 1, 2, and 3, CSH Laboratory Press, Cold Spring Harbor, NY, 1989.

[18] P. Gill, C. Brenner, B. Brinkmann, B. Budowle, A. Carracedo, M.A. Jobling, P. de Knijff, M. Kayser, M. Krawczak, W.R. Mayr, N. Morling, B. Olaisen, V. Pascali, M. Prinz, L. Roewer, P.M. Schneider, A. Sajantila, C. Tyler-Smith, DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, Forensic Sci.Int. 124 (2001) 5-10.

[19] R. Schoske, J.M. Butler, P. Vallone M.C. Kline, M. Prinz, A.J. Redd, M.F. Hammer, Developing Y STR Megaplexes, Proceedings of the 12th International Symposium on Human Identification. (2001) Biloxi, MS, *in press*.

[20] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, Nucleic Acids Res. 29 (2001) 320-322.

[21] S. Rozen, H.J. Skaletsky, Primer3 (primer3_www.cgi v 0.2) Code available at http://www-genome-wi.mit.edu/genome_software/other/primer3.html

[22] J.M. Butler, C.M. Ruitberg, P.M. Vallone, Capillary electrophoresis as a tool for optimization of multiplex PCR reactions, Fresenius.J.Anal.Chem. 369 (2001) 200-205.

[23] J.M. Butler, J.M. Devaney, M.A. Marino, P.M. Vallone, Quality control of PCR primers used in multiplex STR amplification reactions, Forensic Sci.Int. 119 (2001) 87-96.

[24] N. Mathias, M. Bayes, C. Tyler-Smith, Highly informative compound haplotypes for the human Y chromosome, Hum. Mol. Genet. 3 (1994) 115-123.

[25] U. Schmidt, N. Meier, S. Lutz, L. Roewer, A nomenclature for YCAII which is compatible to the ISFG guidelines for Y-STR samples YCAII. Presented at the 19th International Conference on ISFG (2001) Munster, Germany.

[26] C.M Ruitberg and J.M. Butler, New primer sets for Y chromosome and CODIS STR loci, Poster presented at the Eleventh International Symposium on Human Identification (2000) Biloxi, MS.

[27] M.J. Brownstein, J.D. Carpten, J.R. Smith, Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping, Biotechniques. 20 (1996) 1004-1010.

[28] J.M. Clark, Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases, Nucleic Acids Res. 16 (1988) 9677-9686.

[29] P.M. Schneider, S. Meuser, W. Waiyawuth, Y. Seo, C. Rittner, Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations, Forensic Sci.Int. 97 (1998) 61-70.

[30] L. Gusmao, A. Gonzalez-Neira, P. Sanchez-Diz, M.V. Lareu, A. Amorim, A. Carracedo, Alternative primers for DYS391 typing: advantages of their application to forensic genetics, Forensic Sci.Int. 112 (2000) 49-57.

[31] O. Henegariu, N.A. Heerema, S.R. Dlouhy, G.H. Vance, P.H. Vogt, Multiplex PCR: critical parameters and step-by-step protocol, Biotechniques. 23 (1997) 504-511.

[32] A.P. Shuber, V.J. Grondin, K.W. Klinger, A simplified procedure for developing multiplex PCRs, Genome Res. 5 (1995) 488-493.

[33] L. Roewer, M. Kayser, P. Dieltjes, M. Nagy, E. Bakker, M. Krawczak, P.de Knijff, Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations, Hum.Mol.Genet.5 (1996) 1029-1033.

[34] The Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups, Genome Res. 12 (2002) 339-348.

[35] A. Ishii, R.C. Shaler, H.J. Baum, and M. Prinz, Multiplex system for Y-STR testing using fluorescent primers: artifacts caused by female DNA, in: G. Sensabaugh, P.J. Lincoln, B. Olaisen (Eds.), Progress in Forensic Genetics, Vol. 8 Proceedings of the 18th International Conference on ISFH Congress, San Francisco, 17-21 August 1999, Elsevier, Amsterdam, 2000, 457-459.

[36] W. Parson, H. Niederstatter, S. Kochl, M. Steinlechner, B. Berger, When autosomal short tandem repeats fail: optimized primer and reaction design for Y-chromosome short tandem repeat analysis in forensic casework, Croat. Med. J. 42 (2001) 285-287.

[37] B.M. Dupuy, T. Gedde-Dahl, B. Olaisen, DXYS267: DYS393 and its X chromosome counterpart, Forensic Sci. Int. 112 (2000) 111-121.

[38] E. Bosch, A.C. Lee, F. Calafell, E. Arroyo, P. Henneman, P. de Knijff, M.A. Jobling, High resolution Y chromosome typing: 19 Y STRs amplified in three multiplex reactions, Forensic Sci. Int. 125 (2002) 42-51.

[39] S. Furedi, J. Woller, Z. Padar, M. Angyal, Y-STR haplotyping in two Hungarian populations, Int. J. Legal Med. 113 (1999) 38-42.

Table 1. General information on loci used in the Y STR 20plex described here. Information includes the GenBank® accession number for each locus, number of repeats within the GenBank sequence (termed reference allele), repeat motif, and the repeat position within the clone. Some sequences were made reverse and complement (R&C) in order to maintain consistency with previously used forward and reverse primer designations. Repeats motifs for each locus were designated according to the recent International Society of Forensic Genetics (ISFG) guidelines [18].

STR Locus	GenBank® Accession	Reference Allele	Repeat Motif	Repeat Position
DYS19	AC017019 (R&C)	15	$[TAGA]_3TAGG[TAGA]_{12}$	43,291-43,354
DYS385	AC022486 (R&C)	11	GAAA	126,774-126,817
DYS388	AC004810	12	ATT	78,980-79,015
DYS389I	AC004617 (R&C)	12	$[TCTG]_3[TCTA]_9$	62,581-62,628
DYS389II	AC004617 (R&C)	29	$[TCTG]_5[TCTA]_{12}[TCTG]_3$ $[TCTA]_9$	62,581-62,628 62,677-62,744
DYS390	AC011289	24	$[TCTG]_8$ $[TCTA]_{11}$ $[TCTG]_1$ $[TCTA]_4$	69,363-69458
DYS391	AC011302	11	TCTA	31,096-31,139
DYS392	AC011745 (R&C)	13	TAT	52,196-52,234
DYS393	AC006152	12	AGAT	21,115-21,162
DYS426	AC007034	12	GTT	133,611-133,646
DYS437	AC002992	16	$[TCTA]_{10}[TCTG]_2[TCTA]_4$	42,978-43,041
DYS438	AC002531	10	TTTTC	129,828-129,877
DYS439	AC002992	20	$[GATA]_2N_4[GATA]_3N_{14}$ $[GATA]_1N_3[GATA]_1N_7[GATA]_{13}$	91,240-91,347
DYS447	AC005820	23	$[TAATA]_6[TAAAA]_1$ $[TAATA]_9[TAAAA]_1$ $[TAATA]_6$	91,733 – 91,847
DYS448	AC025227	22	$[AGAGAT]_{11}AGAGAGATAG$ $[AGAGAT]_3N_{14}[AGAGAT]_8$	17,246 – 17,401
DYS460 (A7.1)	AC009235 (R&C)	10	ATAG	45,978-46,017
H4	AC011751 (R&C)	12	TAGA	86,623-86,670
YCAII	AC015978	23	CA	79,953-79,998

Table 2. Allele ranges and PCR product sizes. Size ranges given are based on GenBank sequence sizes and take into account adenylation of the PCR products.

STR Locus	Allele Range	New Size Range	Previous Size Range	Reference to locus
DYS19	**10-19**	**233-269 bp**	176-212 bp	[33]
DYS385	**7-23**	**242-306 bp**	241-305 bp	[29]
DYS388	**10-18**	**151-175 bp**	122-146 bp	[2]
DYS389I	**9-17**	**143-175 bp**	142-174 bp	[2]
DYS389II	**26-34**	**263-295 bp**	262-294 bp	[2]
DYS390	**17-28**	**189-233 bp**	188-232 bp	[2]
DYS391	**7-14**	**93-121 bp**	272-300 bp	[2]
DYS392	**6-16**	**290-320 bp**	234-264 bp	[2]
DYS393	**9-16**	**109-133 bp**	108-132 bp	[2]
DYS426	**10-12**	**92-98 bp**	92-98 bp	[11]
DYS437	**14-17**	**186-198 bp**	185-197 bp	[12]
DYS438	**6-13**	**300-335 bp**	202-237 bp	[12]
DYS439	**16-21**	**210-230 bp**	237-257 bp	[12]
DYS447	**22-29**	**207-242 bp**	207-242 bp	[16]
DYS448	**20-26**	**299-335 bp**	289-325 bp	[16]
DYS460	**7-12**	**101-121 bp**	162-182 bp	[14,38]
H4	**8-13**	**122-142 bp**	353-373 bp	[14]
YCAII	**11-24**	**135-161 bp**	135-161 bp	[24]

Table 3. Primer sequences for PCR amplification of loci in the Y STR 20plex. The forward primer (F) is labeled in each case with one of four commercially available fluorescent dyes. A G nucleotide tail was added to the 5'end of the reverse primer (R) in many cases to promote non-template addition. All primer sequences were redesigned compared to previous studies except DYS385 a/b, DYS390, DYS393, DYS437, and DYS426. Predicted primer melting temperatures (T_m) were calculated using a total primer concentration = 0.05 μM and [Na^+] = 50 mM (see ref. [21]).

Locus		Primer Sequences (5'-to-3')	Primer Concentration (μM)	**Primer3** T_m (°C)
DYS19	F	**NED**-ACTACTGAGTTTCTGTTATAGTGTTTTT	1.8	55.0
	R	GTCAATCTCTGCACCTGGAAAT	1.8	60.5
DYS385	F	**VIC**-AGCATGGGTGACAGAGCTA	0.6	56.9
a/b	R	GCCAATTACATAGTCCTCCTTTC	0.6	54.7
DYS388	F	**NED**-GAATTCATGTGAGTTAGCCGTTTAGC	1.8	63.4
	R	GAGGCGGAGCTTTTAGTGAG	1.8	59.2
DYS389	F	**6FAM**-CCAACTCTCATCTGTATTATCTATG	1.3	54.2
I/II	R	GTTATCCCTGAGTAGTAGAAGAATG	1.3	59.0
DYS390	F	**VIC**-TATATTTTACACATTTTTGGGCCC	0.2	57.2
	R	GTGACAGTAAAATGAAAACATTGC	0.2	57.0
DYS391	F	**6FAM**-TTCAATCATACACCCATATCTGTC	0.2	57.9
	R	GATAGAGGGATAGGTAGGCAGGC	0.2	60.0
DYS392	F	**NED**-TAGAGGCAGTCATCGCAGTG	1.8	60.2
	R	GACCTACCAATCCCATTCCTT	1.8	57.3
DYS393	F	**VIC**-GTGGTCTTCTACTTGTGTCAATAC	0.4	54.7
	R	GAACTCAAGTCCAAAAAATGAGG	0.4	57.7
DYS426	F	**VIC**-CTCAAAGTATGAAAGCATGACCA	0.2	59.3
	R	GGTGACAAGACGAGACTTTGTG	0.2	59.8
DYS437	F	**6FAM**-GACTATGGGCGTGAGTGCAT	0.2	61.1
	R	GAGACCCTGTCATTCACAGATGA	0.2	59.6
DYS438	F	**6FAM**-CCAAAATTAGTGGGGAATAGTTG	0.2	58.9
	R	GATCACCCAGGGTCTGGAGTT	0.2	62.6
DYS439	F	**6FAM**-TCGAGTTGTTATGGTTTTAGGTCT	0.2	58.3
	R	GTGGCTTGGAATTCTTTTACCC	0.2	60.3
DYS447	F	**PET**-GGTCACAGCATGGCTTGGTT	0.7	63.4
	R	GGGCTTGCTTTGCGTTATCTCT	0.7	64.0
DYS448	F	**PET**-TGGGAGAGGCAAGGATCCAA	1.1	65.2
	R	GTCATATTTCTGGCCGGTCTGG	1.1	64.3
460	F	**NED**-GAGGAATCTGACACCTCTGACA	0.7	59.3
(A7.1)	R	GTCCATATCATCTATCCTCTGCCTA	0.7	59.1
H4	F	**NED**-ATGCTGAGGAGAATTTCCAA	0.4	57.3
	R	GCTATTCATCCATCTAATCTATCCATT	0.4	56.2
YCAII	F	**VIC**-TGTCAAAATTTAACCCACAATCA	0.9	59.3
a/b	R	GCAGTCTTTCACCATAAGGTTAGC	0.9	57.2

Table 4. Relative positions of primers for each locus compared to previous publications. Primer lengths are in nucleotides (nt).

Locus		Previous primers Distance 3'end to repeat	Primer Length	New primers Distance 3'end to repeat	Primer Length
DYS19	F	55 bp	21 nt	49 bp	28 nt
	R	35 bp	20 nt	89 bp	22 nt
DYS385	F	162 bp	19 nt	*Using published primers*	
a/b	R	8 bp	22 nt		
DYS388	F	19 bp	20 nt	21 bp	26 nt
	R	34 bp	18 nt	53 bp	20 nt
DYS389I	F	3 bp	24 nt	2 bp	25 nt
	R	56 bp	23 nt	55 bp	25 nt
DYS389II	F	7 bp	24 nt	6 bp	25 nt
	R	56 bp	23 nt	55 bp	25 nt
DYS390	F	34 bp	23 nt	*Using published primers*	
	R	39 bp	23 nt		
DYS391	F	14 bp	23 nt	6 bp	24 nt
	R	186 bp	20 nt	11 bp	23 nt
DYS392	F	75 bp	25 nt	225 bp	20 nt
	R	95 bp	20 nt	5 bp	21 nt
DYS393	F	0 bp	24 nt	*Using published primers*	
	R	25 bp	22 nt		
DYS426	F	14 bp	23 nt	*Using published primers*	
	R	2 bp	22 nt		
DYS437	F	10 bp	20 nt	*Using published primers*	
	R	76 bp	22 nt		
DYS438	F	8 bp	21 nt	16 bp	23 nt
	R	122 bp	20 nt	210 bp	21 nt
DYS439	F	73 bp	24 nt	43 bp	24 nt
	R	27 bp	20 nt	23 bp	22 nt
DYS447	F	28 bp	20 nt	28 bp	20 nt
	R	28 bp	20 nt	26 bp	22 nt
DYS448	F	48 bp	22 nt	86 bp	20 nt
	R	46 bp	22 nt	26 bp	22 nt
460	F	72 bp	20 nt	-1 bp	22 nt
(A7.1)	R	16 bp	25 nt	26 bp	25 nt
H4	F	124 bp	27 nt	44 bp	20 nt
	R	116 bp	28 nt	-1 bp	27 nt
YCAII	F	7 bp	21 nt	65 bp	23 nt
a/b	R	63 bp	21 nt	0 bp	24 nt

Table 5. Haplotype results obtained with the Y STR 20plex on the 76 cell lines of the YCC panel. Allele designations were based on sizes rather than comparison to sequenced allelic ladders. YCC1 and YCC54 are female samples and did not generate any allele calls. A triallelic pattern was seen at DYS385 for YCC22. DYS439 failed to amplify for samples YCC29 and YCC30. Origins of the individuals from whom the YCC cell lines were developed are listed at http://ycc.biosci.arizona.edu [34].

Sample	Locus 391	389I	437	19	389II	438	426	393	YCAII	390	385	460	H4	388	439	392	447	448
ycc 1	*female*																	
ycc2	11	14	14	13	30	11	12	13	17-20	23	14-15	11	13	12	19	15	25	22
ycc3	10	14	14	13	31	11	12	15	17-21	21	14-15	11	11	13	18	14	26	22
ycc4	10	13	14	13	30	11	12	13	17-21	24	14-15	11	11	12	20	13	25	21
ycc5	10	13	16	12	28	11	12	11	16	21	15-17	10	8	12	18	13	25	21
ycc6	9	11	14	15	27	11	10	12	17-19	24	10	10	11	12	18	11	27	22
ycc7	9	11	14	15	27	11	10	12	17-19	24	10	10	11	12	18	11	27	22
ycc8	10	12	14	14	28	11	11	13	17-18	25	14-20	11	11	12	18	11	23	22
ycc9	10	14	14	15	30	10	11	15	18	23	13-13	10	11	12	19	11	27	23
ycc 10	10	13	14	15	30	10	11	12	16-20	24	10-18	11	11	12	18	15	27	20
ycc 11	11	12	16	15	27	11	11	13	17-21	24	13-15	10	11	12	19	13	26	22
ycc12	10	13	14	13	30	11	12	13	17-21	24	14-15	10	11	13	19	15	25	23
ycc13	10	13	14	13	30	11	12	13	17-21	24	14-15	10	11	13	19	15	25	23
ycc14	11	13	14	13	30	11	12	13	17-23	24	19	11	11	12	18	14	25	22
ycc15	10	13	14	13	30	11	12	13	17-21	24	13-17	11	12	12	18	14	24	21
ycc16	10	13	14	13	30	11	12	13	17-21	24	13-17	11	12	12	18	14	24	21
ycc17	10	13	14	13	30	11	12	13	17-19	25	14-16	10	12	12	20	14	24	22
ycc18	10	13	15	13	29	11	12	13	15-21	23	14-18	11	12	12	20	14	24	22
ycc19	10	12	14	17	28	11	12	13	18-20	19	14-20	11	12	11	17	9	23	23
ycc21	10	14	14	16	26	10	11	12	19-20	24	12-14	10	12	12	19	11	25	24
ycc22	10	14	16	12	29	11	12	11	16	21	15-16-17	10	8	12	18	13	25	21
ycc23	10	13	14	16	28	10	11	13	19-21	23	13	10	11	13	18	11	29	23
ycc24	10	12	16	15	31	10	11	14	19	22	15	10	11	12	19	10	26	24
ycc25	10	13	14	13	30	11	12	13	17-21	25	14	10	11	12	19	14	24	22
ycc26	11	13	15	14	29	12	12	13	17-20	23	11-14	11	12	12	18	13	24	22
ycc27	11	13	15	15	28	12	12	13	17-21	24	11-14	11	12	12	18	13	24	20
ycc28	10	14	15	15	26	11	11	12	19-20	24	11-14	11	12	12	17	11	25	25
ycc29	9	14	15	16	30	10	10	13	19-20	20	15-16	11	11	12		11	27	25
ycc30	9	14	15	16	30	10	10	13	19-20	20	15-16	11	11	12		11	27	25
ycc31	11	13	14	15	30	11	11	14	17-19	21	16	11	11	12	19	12	25	24
ycc32	11	10	14	13	27	10	11	14	17-20	24	16-17	11	11	12	17	11	27	23
ycc33	10	14	14	15	31	11	11	13	17-17	21	15-19	10	12	12	18	11	25	24
ycc34	10	13	16	13	28	10	12	11	16	21	16-17	10	8	12	19	13	24	23
ycc35	11	13	16	13	29	10	12	11	16	22	16-17	10	8	12	19	13	23	23
ycc36	10	12	14	15	30	11	11	14	17-19	21	13-17	10	12	12	18	11	25	24
ycc37	10	12	14	14	28	11	11	13	17-18	25	14-20	11	11	12	18	11	23	22
ycc38	11	12	14	17	28	11	12	13	18-19	18	14-18	11	11	11	17	10	24	22
ycc39	11	13	15	15	25	11	11	12	19-20	24	11-15	11	12	12	17	11	25	25
ycc40	10	13	13	15	30	11	11	14	17	21	17-19	10	11	12	18	11	27	24
ycc42	10	13	14	15	31	10	11	13	19-20	24	11	11	13	10	20	11	27	26
ycc43	9	13	14	16	30	11	11	15	17-19	21	17-20	11	11	12	19	11	27	24
ycc44	11	13	14	16	30	11	11	14	17-19	21	14-16	10	11	12	20	11	25	24
ycc45	11	13	14	15	30	11	11	13	17	21	17-18	11	12	12	19	11	25	24

ycc47	**11**	**14**	**14**	**14**	**31**	**11**	**11**	**14**	**16-18**	**23**	**11-13**	**11**	**12**	**12**	**17**	**16**	**24**	**22**
ycc48	**11**	**14**	**14**	**14**	**32**	**11**	**11**	**14**	**16-18**	**23**	**11-13**	**11**	**12**	**12**	**17**	**16**	**24**	**22**
ycc49	**11**	**14**	**14**	**14**	**31**	**11**	**11**	**14**	**16-18**	**23**	**11-13**	**11**	**12**	**12**	**17**	**15**	**24**	**22**
ycc50	**11**	**14**	**14**	**14**	**31**	**11**	**11**	**14**	**16-18**	**23**	**11-13**	**11**	**12**	**12**	**17**	**15**	**24**	**22**
ycc51	**11**	**14**	**14**	**14**	**31**	**11**	**11**	**14**	**16-20**	**23**	**11-13**	**11**	**12**	**12**	**17**	**16**	**25**	**22**
ycc52	**10**	**12**	**16**	**15**	**29**	**10**	**11**	**14**	**18**	**23**	**13-14**	**11**	**12**	**12**	**18**	**11**	**23**	**23**
ycc53	**10**	**13**	**15**	**15**	**31**	**10**	**11**	**14**	**18**	**23**	**13-14**	**12**	**12**	**12**	**18**	**12**	**23**	**23**
ycc54	*female*																	
ycc55	**10**	**12**	**16**	**15**	**30**	**10**	**11**	**14**	**18**	**23**	**13-14**	**10**	**12**	**12**	**18**	**11**	**23**	**23**
ycc56	**10**	**14**	**15**	**14**	**31**	**9**	**11**	**12**	**18**	**22**	**13-14**	**10**	**11**	**15**	**18**	**11**	**23**	**24**
ycc57	**10**	**12**	**15**	**15**	**28**	**10**	**11**	**12**	**17-20**	**23**	**12-16**	**8**	**12**	**12**	**18**	**12**	**24**	**22**
ycc58	**10**	**13**	**14**	**15**	**29**	**9**	**11**	**12**	**17-19**	**22**	**16**	**11**	**11**	**13**	**18**	**11**	**24**	**21**
ycc59	**10**	**14**	**16**	**14**	**30**	**10**	**11**	**12**	**19-20**	**25**	**14**	**10**	**11**	**15**	**18**	**11**	**25**	**23**
ycc60	**10**	**12**	**15**	**14**	**28**	**9**	**11**	**13**	**17-20**	**23**	**13-18**	**10**	**11**	**17**	**19**	**11**	**26**	**24**
ycc61	**11**	**13**	**14**	**15**	**29**	**10**	**11**	**15**	**17-19**	**23**	**15-16**	**11**	**11**	**13**	**19**	**12**	**24**	**23**
ycc62	**11**	**13**	**15**	**14**	**29**	**12**	**12**	**13**	**17-21**	**23**	**11-14**	**11**	**12**	**12**	**19**	**13**	**25**	**22**
ycc63	**10**	**12**	**16**	**15**	**29**	**10**	**11**	**13**	**17-19**	**22**	**13-14**	**10**	**11**	**14**	**18**	**11**	**22**	**24**
ycc64	**12**	**13**	**14**	**14**	**28**	**12**	**12**	**13**	**17-21**	**24**	**11-15**	**11**	**11**	**12**	**20**	**13**	**25**	**23**
ycc65	**10**	**13**	**14**	**15**	**30**	**11**	**11**	**14**	**17**	**21**	**16-18**	**10**	**13**	**12**	**18**	**11**	**25**	**24**
ycc66	**10**	**12**	**14**	**15**	**28**	**10**	**11**	**13**	**18**	**23**	**12-13**	**11**	**12**	**12**	**19**	**14**	**25**	**21**
ycc67	**10**	**12**	**14**	**15**	**27**	**10**	**11**	**13**	**17-21**	**23**	**13-13**	**10**	**12**	**12**	**19**	**14**	**25**	**21**
ycc68	**10**	**13**	**15**	**15**	**28**	**10**	**11**	**12**	**16-20**	**24**	**13-23**	**9**	**11**	**12**	**20**	**14**	**25**	**22**
ycc69	**11**	**13**	**14**	**15**	**30**	**10**	**11**	**14**	**17-20**	**25**	**16-20**	**11**	**11**	**12**	**19**	**13**	**27**	**21**
ycc70	**10**	**13**	**14**	**17**	**29**	**11**	**12**	**13**	**17-21**	**24**	**9-14**	**11**	**12**	**12**	**18**	**11**	**23**	**23**
ycc71	**10**	**13**	**15**	**14**	**29**	**12**	**13**	**13**	**17-21**	**24**	**12-15**	**10**	**11**	**12**	**19**	**13**	**26**	**22**
ycc72	**11**	**13**	**15**	**15**	**32**	**10**	**11**	**13**	**19**	**24**	**13-15**	**10**	**11**	**13**	**20**	**11**	**25**	**23**
ycc74	**11**	**13**	**15**	**16**	**32**	**10**	**11**	**13**	**19**	**24**	**14-15**	**10**	**11**	**13**	**21**	**11**	**25**	**23**
ycc76	**10**	**13**	**14**	**17**	**30**	**10**	**11**	**13**	**17-19**	**26**	**14-18**	**10**	**12**	**12**	**18**	**11**	**25**	**21**
ycc77	**10**	**14**	**14**	**13**	**30**	**10**	**11**	**13**	**16-17**	**23**	**12-13**	**11**	**12**	**12**	**18**	**16**	**28**	**22**
ycc78	**10**	**12**	**15**	**15**	**28**	**10**	**11**	**13**	**18-20**	**23**	**11-12**	**10**	**13**	**12**	**19**	**12**	**24**	**22**
ycc79	**10**	**12**	**15**	**14**	**28**	**10**	**11**	**13**	**18**	**24**	**13-15**	**10**	**10**	**12**	**18**	**12**	**23**	**23**
ycc80	**10**	**12**	**16**	**16**	**31**	**10**	**11**	**14**	**19**	**22**	**15**	**11**	**11**	**12**	**19**	**10**	**26**	**24**
ycc81	**10**	**10**	**14**	**15**	**27**	**11**	**12**	**14**	**17-21**	**25**	**11-14**	**10**	**11**	**12**	**17**	**12**	**24**	**23**

Table 6. PowerPlex® 16 genotypes for six pairs of YCC samples that could not be distinguished from one another by their Y STR 20plex haplotypes. In each case, samples are from different individuals that are likely male relatives to the other member of the pair due to the high number of shared autosomal STR alleles (bold).

STR Markers	**ycc6**	**ycc7**	**ycc12**	**ycc13**	**ycc15**	**ycc16**	**ycc29**	**ycc30**	**ycc49**	**ycc50**	**ycc8**	**ycc37**
D3S1358	15,**15**	14,**15**	16,**16**	15,**16**	15,**15**	15,**15**	15,**16**	**16**,17	15,15	16,17	**15**,17	**15**,15
TH01	**7**,**7**	**7**,**7**	**6**,6	**6**,7	6,**7**	7,**7**	**8**,8	**8**,10	**7**,9	6,**7**	8,9	6,6
D21S11	**28**,34	**28**,36	**29**,**32.2**	**29**,**32.2**	32.3,**33.2**	31.2,**33.2**	**28**,30	**28**,29	28,30	29,32.2	26,31	29,33.2
D18S51	**15**,19	**15**,15	**14**,18	**14**,15	14,17	15,18	**12**,18	**12**,24	**17**,21	**17**,17	**17,19**	**17,19**
Penta E	8,10	12,12	**8**,15	**8**,20	**12**,12	**12**,16	5,12	10,11	11,13	10,14	8,10	5,7
D5S818	**8**,12	**8**,10	**10**,11	7,**10**	**7,11**	**7,11**	12,12	11,14	10,**11**	**11**,12	13,**13**	11,**13**
D13S317	**12**,13	**12**,12	**9**,10	**9**,13	**9**,9	**9**,13	11,**12**	12,**12**	**11**,13	**11**,11	11,**12**	**12**,13
D7S820	10,**10**	9,**10**	11,11	10,12	**11,11**	**11,11**	10,10	9,11	**8**,12	**8**,8	**8**,8	**8**,12
D16S539	**9**,9	**9**,10	9,10	12,12	**10,12**	**10,12**	10,12	9,11	**9**,11	**9**,12	10,12	9,9
CSF1PO	**10**,**12**	**10**,**12**	**11,11**	**11,11**	**10**,11	**10**,10	**6**,6	**6**,11	11,**12**	12,**12**	**10**,14	7,**10**
Penta D	**2.2**,8	**2.2**,9	**9,10**	**9,10**	9,**13**	11,**13**	2.2,13	9,10	12,12	8,9	**2.2,8**	**2.2,8**
VWA	**15**,**16**	**15**,**16**	**16**,16	**16**,18	**16**,16	**16**,18	18,**19**	16,**19**	**18**,19	17,**18**	15,**16**	**16**,18
D8S1179	14,14	11,15	**10**,10	**10**,13	14,15	10,13	11,**15**	13,**15**	**13**,14	10,**13**	12,**14**	**14**,15
TPOX	**9,9**	**9,9**	**8**,12	**8**,11	12,12	8,8	8,**10**	9,**10**	**11,11**	**11,11**	9,**11**	8,**11**
FGA	**21**,22	**21**,24	**21**,22	19,**21**	25,**25**	21,**25**	25,25	20,24	25,25	24,24	20,24	22,25
Amelogenin	X,Y	X,Y	X,Y	X,Y	X,Y	X,Y	X,Y	X,Y	X,Y	X,Y	X,Y	X,Y

Figure 1. Schematic of PCR product size ranges produced with the known allele size ranges (see Table 2) for the loci in the Y STR 20plex. Marker names have been abbreviated (e.g., DYS391 is listed as 391). The locus A7.1 was recently designated DYS460 [38].

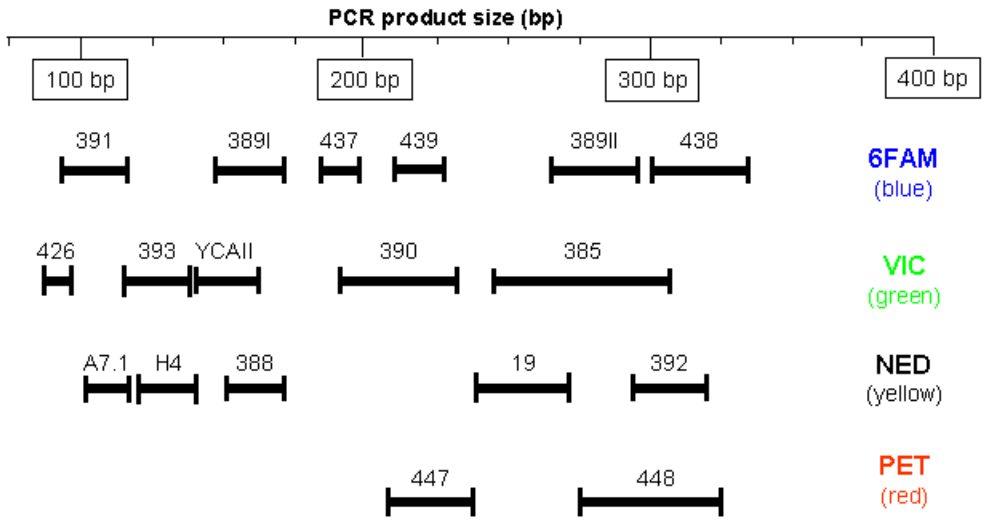


Figure 2. Alignment of top strands from DYS391 X and Y homologous sequences. The Y sequence comes from GenBank accession AC011302 while the X sequence is from AF055717. The X sequence is missing 29 nucleotides (indicated by dashes) compared to the Y sequence for DYS391. Sequence differences between the X and Y homologs are boxed. The primers indicated by the dotted arrows are the commonly used ones that are listed in the Genome Database (GDB). The Gusmao R2 primer (Ref. [30]) targeted a region containing four nucleotide differences between the X and Y sequences. The primers shown with the solid arrows are the new ones described here that contain a single base difference in the forward and 13 base differences (circled regions) in the reverse primer relative to the X sequence to make them more Y chromosome specific.

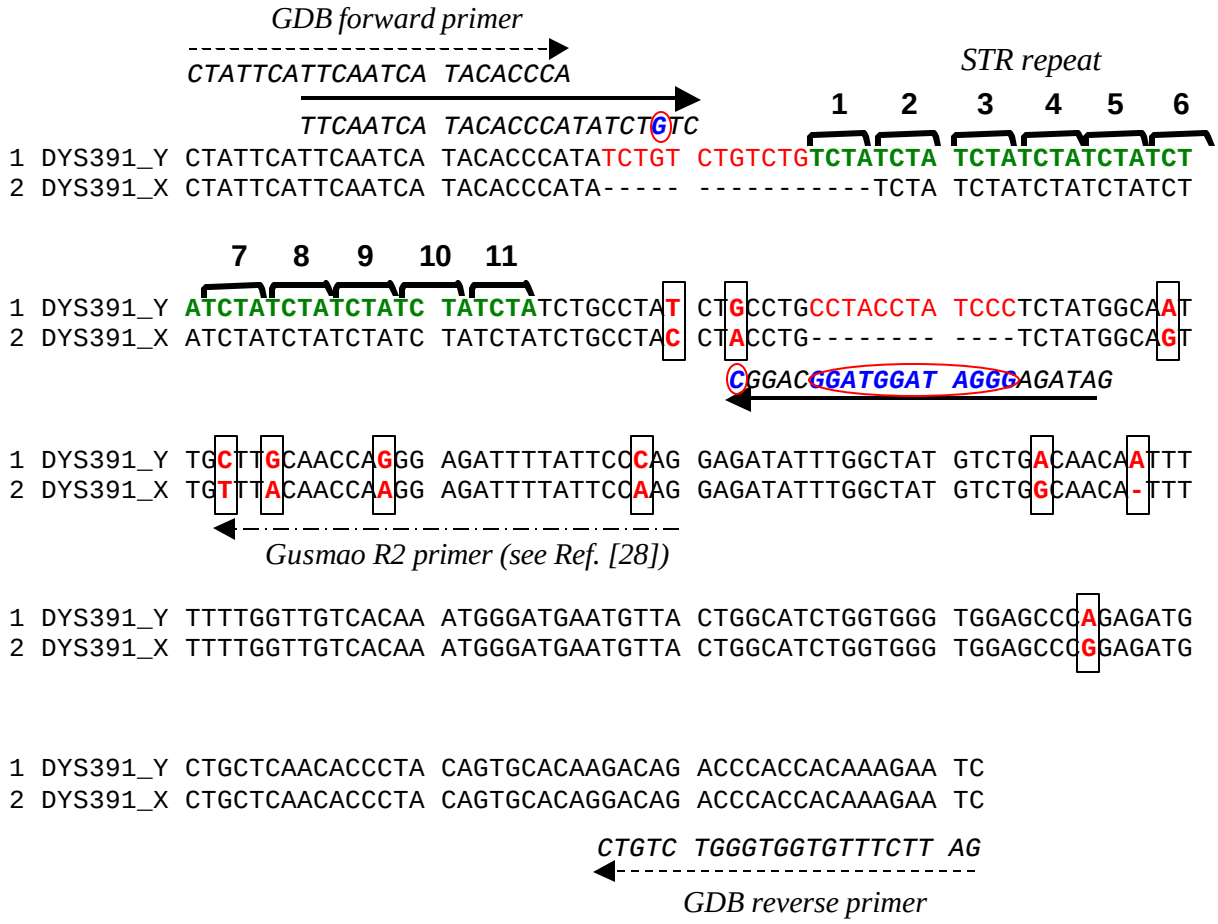
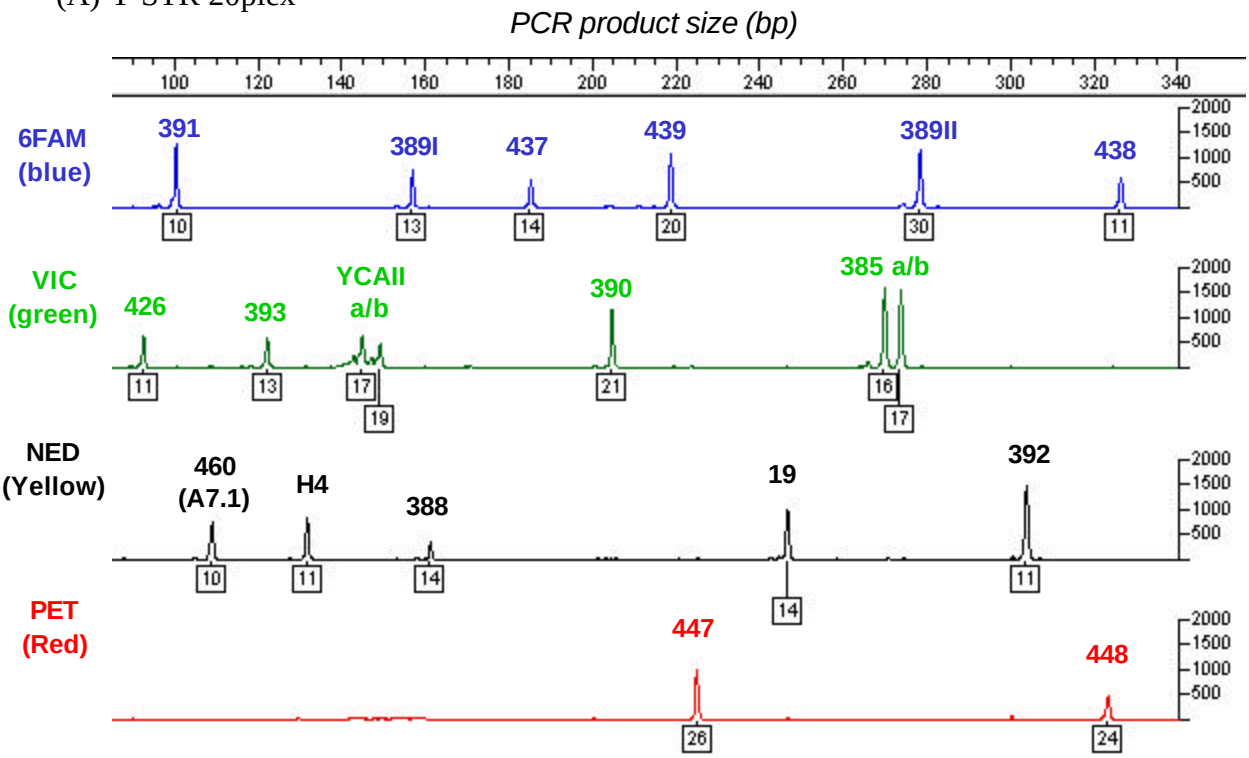


Figure 3. Genotyper® results for the same male DNA sample obtained using (A) the Y STR 20plex and (B) the Y-PLEX™ 6 kit from ReliaGene. Typing results between the two multiplexes were identical on the 6 loci (7 typing results) that overlapped.

(A) Y STR 20plex



(B) Y-PLEX™ 6 Kit

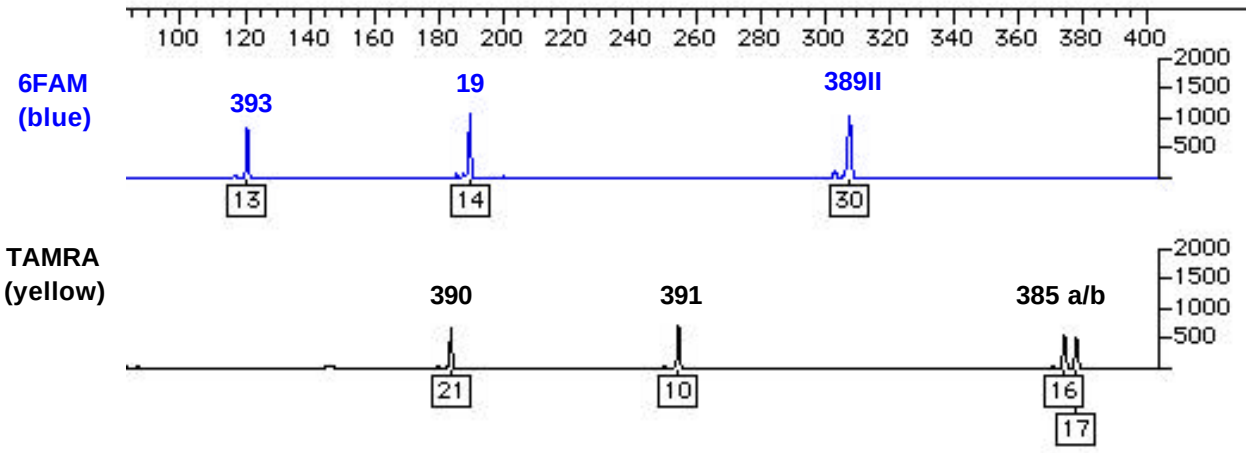


Figure 4. GeneScan® results from the same male DNA sample amplified with the Y STR 20plex (top) compared to an 11plex (bottom) that contains only the European extended haplotype loci.